

Tilburg University

Faking on direct, indirect, and behavioural measures of spider fear

Langner, O.; Ouwens, M.A.; Muskens, M.; Trumpf, J.; Becker, E. S.; Rinck, M.

Published in:
Cognition and Emotion

DOI:
[10.1080/02699930802652412](https://doi.org/10.1080/02699930802652412)

Publication date:
2010

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Langner, O., Ouwens, M. A., Muskens, M., Trumpf, J., Becker, E. S., & Rinck, M. (2010). Faking on direct, indirect, and behavioural measures of spider fear: Can you get away with it? *Cognition and Emotion*, 24, 549-558. <https://doi.org/10.1080/02699930802652412>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

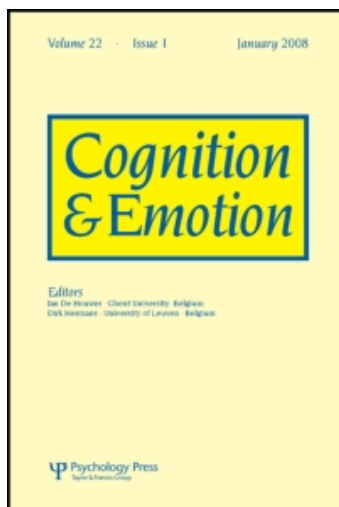
This article was downloaded by: [Tilburg University]

On: 9 September 2009

Access details: Access Details: [subscription number 907217989]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Cognition & Emotion

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713682755>

Faking on direct, indirect, and behavioural measures of spider fear: Can you get away with it?

Oliver Langner ^a; Machteld Ouwens ^a; Marjolein Muskens ^a; Julia Trumpf ^a; Eni S. Becker ^a; Mike Rinck ^a

^a Radboud University, Nijmegen, The Netherlands

First Published on: 29 January 2009

To cite this Article Langner, Oliver, Ouwens, Machteld, Muskens, Marjolein, Trumpf, Julia, Becker, Eni S. and Rinck, Mike(2009)'Faking on direct, indirect, and behavioural measures of spider fear: Can you get away with it?',Cognition & Emotion,99999:1,

To link to this Article: DOI: 10.1080/02699930802652412

URL: <http://dx.doi.org/10.1080/02699930802652412>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

BRIEF REPORT

Faking on direct, indirect, and behavioural measures of spider fear: Can you get away with it?

Oliver Langner, Machteld Ouwens, Marjolein Muskens,
Julia Trumpf, Eni S. Becker, and Mike Rinck

Radboud University, Nijmegen, The Netherlands

We tested direct, indirect, and behavioural measures of fear of spiders under neutral instructions, and when participants were asked to fake high and low fear of spiders. Our findings indicate that the Approach Avoidance Task (AAT) was the only measure that could be faked in one of the faking conditions only. We also assessed how easily faked results could be detected on each measure for different diagnostic criteria. The direct and behavioural measures showed good performance for all criteria. The AAT performed comparably only for a conservative criterion, when detecting fakers is less important than correctly labelling non-fakers.

Keywords: Indirect measures; AAT; Direct measures; Faking; Fake-detection.

In the measurement of anxiety and other psychiatric phenomena, direct measures such as questionnaires are often criticised for their susceptibility to response bias and faking. Critically, direct measures rely on both the accessibility of information and the participants' willingness to accurately report it. As an alternative, several researchers have developed indirect measures. Instead of asking directly, these measures assess responses like reaction times, which are indirectly related to, but significantly influenced by, the measured concept.¹ Because the relationship between task and influencing concept is not obvious, these measures are generally assumed to suffer

¹ By indirect measures we refer mainly to reaction time tasks, as those are widely used. But indirect measures of other variables than latencies exist.

Correspondence should be addressed to: Oliver Langner, Behavioural Science Institute, Montessorilaan 3, NL-6525 HR Nijmegen, The Netherlands. E-mail: o.langner@psych.ru.nl

This study was funded by a research grant of the Behavioural Science Institute of the Radboud University Nijmegen.

less from susceptibility problems (Fazio & Olson, 2003). So far, however, there exists little evidence supporting this claim.

Studies testing the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) have delivered equivocal results: Whereas some found that it was less prone to faking than explicit measures (Schnabel, Banse, & Asendorpf, 2006; Steffens, 2004), or could be faked only after giving specific instructions (Kim, 2003), others found that prior task experience is sufficient to invent faking strategies (Fiedler & Bluemke, 2005; Steffens, 2004). Further, faking the IAT might be easier if implicit attitudes are not deeply elaborated (De Houwer, Beckers, & Moors, 2007).

Faking has also been tested for other indirect measures: Recent studies showed affective priming to be susceptible to faking and strategic effects (Klauer & Teige-Mocigemba, 2007; Teige-Mocigemba & Klauer, 2008). More promising results have been found for the non-latency-based Conditional Reasoning Test, as faking occurred only when the assessment purpose was disclosed (LeBreton, Barksdale, Robin, & James, 2007). Clearly, further testing with other measures is needed to determine whether indirect measures can be considered a viable alternative to direct measures, and if the claim about their lower faking susceptibility is justified.

A shortcoming of most faking studies is that direct and indirect measures were rarely directly compared under faking conditions to ascertain how each is affected by intentional deception (Schnabel et al., 2006). Further, employing direct and indirect measures along with observable behavioural responses can provide information regarding the controllability of various aspects of the underlying processes. In this study, fear of spiders provided a practical framework to examine these questions, because established direct, indirect, and behavioural measures with sufficient variance within the population are available. We used the Spider Anxiety Screening (SAS; Rinck et al., 2002) and the Fear of Spider Questionnaire (FSQ; Szymanski & O'Donohue, 1995) as direct measures, a Behavioural Approach Test (BAT) as a behavioural measure, and the Approach Avoidance Task (AAT) as an indirect measure of fear for spiders. The AAT involves moving a joystick towards or away from oneself in response to a stimulus. It is based on findings that participants respond faster when pushing aversive stimuli away or when pulling pleasant stimuli towards themselves (Rinck & Becker, 2007; Solarz, 1960). Results for a similar joystick-based procedure (IAP; Schnabel et al., 2006) were already promising in that it was found to be more robust against faking than direct measures, and as robust as the IAT.

We asked participants to perform all tests under a neutral and two faking conditions. In the neutral condition, participants received the standard instructions for each test. In the fake-high and fake-low conditions, we instructed participants to adjust their responses to appear either highly fearful or not at all fearful of spiders. These rather unspecific instructions, in

our view, reflect well the situation faced by someone with faking intentions, as most participants would not have extensive prior knowledge regarding the task and thus can only rely on experience gained during the task.

We additionally tested in which direction (fake-low or fake-high anxiety) faking was easier to accomplish on each measure. This is particularly important as fakeability may not be symmetric in both directions on all measures. For example, while responding more slowly on a reaction time (RT) measure may be possible, responding more quickly may prove difficult.

Besides testing fakeability, it is important to know how easily faked results can be detected. In general, faking is harmful only if faked results can not be identified. In this sense, a good measure is either hard to fake or makes it easy to identify fakers. Attempts to assess successful faking have been made for the IAT (Fiedler & Bluemke, 2005), but faked results could not reliably be detected. Here, we determined cutoff values for each measure to categorise data as faked for three different diagnostic criteria, and we calculated d' values to quantify how well fakers could be identified using this strategy.

METHOD

Participants

A total of 59 students (50 female) from Radboud University Nijmegen participated in the study, with mean age of 22.1 years ($SD = 5.1$). Participants completed the SAS prior to the experiment. Pre-test-SAS scores ranged from 0 to 23 with mean 9.8 ($SD = 7.2$), nearly spanning the possible range of 0 to 24. All participants had normal or corrected-to-normal vision and received 9 Euro or course credit for participation.

Questionnaires

We used two established direct measures of fear of spiders: the SAS (Rinck et al., 2002) and the FSQ (Szymanski & O'Donohue, 1995). The SAS has 4 and the FSQ 18 items, both with scales ranging from 0 to 6. We used the Dutch FSQ (Muris & Merckelbach, 1996) and created a Dutch version of the SAS via back-translation. All items were successively presented on a monitor and responses were recorded by the computer.

AAT stimuli

AAT stimuli consisted of eight spider and eight butterfly pictures serving as negatively and positively valenced stimuli, and a white frame with black background as neutral stimulus. To reflect natural variability, we chose pictures of butterflies and spiders that varied in size, colour, and shape. For

each picture, high and wide format versions were created, with image dimensions of 400×320 and 320×400 pixels, respectively. To produce a zoom effect while moving the joystick, we used similar image sizes as Rinck and Becker (2007), varying between 80×100 and 560×700 pixels.

Reliabilities

The direct measures have good psychometric properties, with Cronbach's alphas of .96 and .92 for the FSQ and SAS, respectively; and with test-retest reliabilities between .91 and .95 for the FSQ and .88 for the SAS (Muris & Merckelbach, 1996; Rinck et al., 2002). Split-half reliabilities for the AAT were generally lower, between .71 and .80 (Rinck & Becker, 2007).

Procedure

The experiment had a neutral and two faking conditions. In the neutral condition, participants performed all tests under standard instructions. In the fake-high (fake-low) condition, we instructed participants as follows:

Please try to behave like someone who is *highly afraid* (not afraid) of spiders. Even if you are in reality *not afraid* (afraid) of spiders, please try to fake your performance on the tasks so that someone analysing your results will believe that you are *highly afraid of spiders* (not afraid of spiders at all).

Task order was counterbalanced across subjects but remained constant throughout all experimental conditions. All participants started with the neutral condition, while the order of faking conditions was counterbalanced across participants. This ensured that the neutral condition was not influenced by the faking instructions, and gave participants an opportunity to gain experience with all the measures, which they could use to generate faking strategies (Steffens, 2004). The experiment took about 50 minutes.

The AAT consisted of pushing or pulling a joystick in response to the picture format. Half of the participants pushed high-format and pulled wide-format pictures, while the other half received opposite instructions. Picture content was irrelevant for this task. We instructed participants to respond quickly without committing too many errors. Each AAT had 160 trials and took 10 minutes. The 32 high- and wide-format spider and butterfly pictures were each shown 4 times, yielding 128 trials. Additionally, 16 high- and 16 wide-format empty-frame stimuli were presented. Stimulus order was randomised with the restriction that identical pictures did not appear consecutively. In the neutral condition, the AAT started with 20 practice trials containing empty frame stimuli. Participants started each trial by moving the joystick to the centre position and pressing a button. Then a picture appeared on the screen, and participants had to indicate by a joystick

movement whether the image was in high or wide format. To strengthen the impression of pushing the image away or pulling it closer, the displayed images shrank or grew corresponding to the joystick movement. Moving the joystick by 30° in either direction made the picture disappear. We measured reaction time (RT) as time between stimulus onset and offset.

The BAT was conducted in a separate room, in which a terrarium with a living tarantula was located on a table 4 m from the door. We took participants to the closed room and instructed them to open the door and approach the spider as quickly and closely as possible. Participants could stop whenever they wanted. The time taken to approach the spider was registered, as was the remaining distance to the spider.

RESULTS

Reported *p*-values are based on Greenhouse–Geisser corrected degrees of freedom (*dfs*) where appropriate. For readability, uncorrected *dfs* are reported along with the *df*-correction parameters.

Fakeability of measures

Self-report measures. For both questionnaires, we found strong effects of faking condition. Sum scores were significantly lower in the fake-low condition and significantly higher in the fake-high condition compared to the neutral condition. Average sum scores (*SDs*) were 2.4 (10.8), 19.2 (23.7), and 98.8 (12.1) for the FSQ; and 0.9 (3.5), 8.8 (7.4), and 22.9 (2.1) for the SAS in the fake-low, neutral, and fake-high conditions, respectively. An ANCOVA with faking condition as repeated-measures factor and pre-test-SAS score as covariate yielded significant main effects of faking condition for the FSQ, $F(2, 114) = 635.3$, $p < .01$, $\eta_p^2 = .92$, $\varepsilon = .89$; and for the SAS, $F(2, 114) = 374.8$, $p < .01$, $\eta_p^2 = .87$, $\varepsilon = .62$. Pairwise comparisons between all faking conditions were significant for both questionnaires, all $ts(58) > 5.6$, $p < .01$.

Behavioural measure. We calculated the average approach speed on the BAT in centimetres per second. When participants refused to enter the room, which happened only in the fake-high condition, their speed was set to zero. As expected, participants adjusted their speed corresponding to faking conditions: They approached the spider significantly faster in the fake-low and significantly slower in the fake-high condition, compared to the neutral condition. Average approach speeds (*SDs*) were 70.9 (31.4), 54.7 (33.0), and 9.6 (11.6) cm/s in the fake-low, neutral, and fake-high conditions, respectively. The ANCOVA with factors faking condition and pre-test-SAS score showed a significant main effect of faking condition, $F(2, 114) = 85.9$,

TABLE 1
Mean percentage of errors and mean reaction times (RTs, in milliseconds) of the AAT.
Standard deviations are given in parentheses

Picture type	Instructions					
	Fake-low		Neutral		Fake-high	
	Pull	Push	Pull	Push	Pull	Push
RTs						
Butterfly	743 (131)	736 (149)	801 (135)	803 (158)	872 (186)	846 (202)
Empty	712 (121)	692 (137)	766 (139)	769 (143)	794 (134)	762 (151)
Spider	726 (134)	749 (155)	784 (133)	815 (158)	1100 (514)	788 (206)
Error rates						
Butterfly	2.6 (3.5)	1.7 (3.5)	2.8 (3.4)	2.4 (3.4)	4.0 (5.8)	5.6 (7.6)
Empty	1.1 (2.6)	1.3 (2.8)	2.3 (3.4)	1.6 (3.2)	1.7 (4.2)	2.1 (4.1)
Spider	1.6 (2.9)	2.5 (3.9)	2.8 (3.7)	2.2 (3.1)	42.0 (41.0)	2.0 (3.4)

$p < .01$, $\eta_p^2 = .60$, $\varepsilon = .92$. All pairwise comparisons between faking conditions were significant, all $ts(58) \geq 2.95$, $p < .02$.

Indirect measure: RTs. We determined median RTs for each participant and AAT condition.² A $3 \times 3 \times 2$ (Picture Type \times Faking Condition \times Movement) repeated-measures ANCOVA with pre-test-SAS score as covariate showed significant main effects and interactions for all non-SAS-related effects, $F \geq 7.97$, $p < .01$, $.32 \leq \varepsilon \leq .74$ for all effects. Post hoc analyses revealed an obvious strategy that participants had used in the fake-high condition: Participants responded much more slowly when they had to pull spider pictures closer (see Table 1), yielding a significant Faking Condition \times Movement interaction when analysing spider pictures only, $F(2, 104) = 26.9$, $p < .01$, $\eta_p^2 = .34$, $\varepsilon = .53$. This was confirmed by a significant contrast for spider pictures, comparing RTs of pull responses in the fake-high condition to the average RTs of all other conditions, $t(53) = 4.69$, $p < .01$ (one-tailed), $r_{\text{contrast}} = .54$.

Indirect measure: Errors. We calculated the percentage of errors for each subject and AAT condition. Prior to analyses, percentages were arcsin-

² Usually AAT data are analysed as AAT effects: For each condition, median pull RTs are subtracted from the corresponding push RTs. Positive (negative) AAT effects indicate stronger approach (avoidance) tendencies towards the stimulus. As faking appeared systematically on RT level, the analysis of AAT effects is omitted.

transformed, although for readability non-transformed percentages are reported (see Table 1).³

A $3 \times 3 \times 2$ (Picture Type \times Faking Condition \times Movement) repeated-measures ANCOVA with pre-test-SAS score as covariate was calculated. All non-SAS-related main effects and interactions were significant, $F \geq 28.4$, $p < .01$, $\eta_p^2 \geq .34$, $.29 \leq \varepsilon \leq .59$ for all effects. Post hoc tests confirmed an obvious pattern: Participants made more errors in the fake-high condition for pulling spider pictures towards themselves, $t(58) > 6.89$, $p < .01$ (for pairwise comparisons, see Table 1). Error rates for this stimulus–movement combination were 42%, 2.8%, and 1.6% in the fake-high, neutral, and fake-low conditions, respectively. Obviously, participants often deliberately pushed spider pictures in the fake-high condition, despite clear instructions to pull. A weaker, but similar effect occurred for butterfly pictures: Participants made more errors in the fake-high condition compared to the other faking conditions, when pushing butterfly pictures away from themselves, $t(58) > 2.98$, $p < .01$. Error rates for pushing butterflies were 5.6%, 2.4%, and 1.8% in the fake-high, neutral, and fake-low conditions, respectively. Participants deliberately pulled butterfly pictures in the fake-high condition, despite instructions to push.

Detectability of faked results

To assess the detectability of faking, data from the fake-low and fake-high conditions were compared to those from the neutral condition. For each measure, we constructed empirical ROC curves from the hit and false alarm rates of all possible cutoff values, and fitted the best ROC curve (Macmillan & Creelman, 2004). The hit rate is the proportion of fake-high or fake-low responses correctly identified as faked, whereas the false alarm rate is the proportion of neutral responses incorrectly labelled as faked. As most ROC curves were asymmetric due to different variances in the faked and non-faked data, the average detection performance is best characterised by the measure d_a (Macmillan & Creelman, 2004). Values for d_a of 0, 1, 2, and 3 correspond to a proportion of correct identifications of .50, .69, .84, and .93, respectively.

We additionally report cutoff values for three diagnostic criteria: a conservative criterion (correct identification of non-fakers twice as important as detection of fakers), an unbiased criterion (both goals equally important), and a lenient criterion (detection of fakers twice as important as correctly identifying non-fakers), corresponding to slopes of 2, 1, and 0.5 on the ROC curve, respectively (Swets, Dawes, & Monahan, 2000). In Table 2,

³ We used the arcsine transformation of the form $Y = 2\arcsin\sqrt{p}$, which is standardly used to correct for the inherent dependence of means and variances in proportion data (Winer, 1971).

TABLE 2
Cutoff values, hit rates (H), false alarm (FA) rates, and d' values for detection of faked results under fake-high and fake-low conditions.
Results for three diagnostic criteria and either for the study sample only or for data from other studies included

Test	Neutral vs. fake-high							Neutral vs. fake-low						
	Study sample				Including other studies			Study sample				Including other studies		
	Cutoff	H	FA	d'	H	FA	d'	Cutoff	H	FA	d'	H	FA	d'
Conservative criterion														
FSQ	≥ 72	.93	.03	3.36	.93	.13	2.60	²						
SAS	≥ 21	.90	.05	2.93	.90	.18	2.20	≤ 0	.78	.08	2.17	.78	.17	1.72
BAT	≤ 21 cm/s	.88	.05	2.82	.88	.15	2.21	≥ 81 cm/s	.27	.08	0.79	.27	.14	0.47
AAT RT	≥ 1063 ms	.36	.02	1.7	.36	.02	1.70	¹						
AAT error	$\geq 13\%$.58	.02	2.26	.58	.01	2.53	¹						
AAT comb.		.76	.03	2.59	.76	.03	2.59	¹						
Unbiased criterion														
FSQ	≥ 70	.95	.03	3.53	.95	.15	2.68	≤ 0	.73	.20	1.45	.73	.22	1.39
SAS	≥ 20	.93	.10	2.76	.93	.24	2.18	²						
BAT	≤ 24 cm/s	.93	.10	2.76	.93	.21	2.28	≥ 53 cm/s	.73	.46	0.71	.73	.48	0.66
AAT RT	≥ 962 ms	.42	.12	0.97	.42	.05	1.44	¹						
AAT error	$\geq 4\%$.68	.19	1.35	.68	.14	1.55	¹						
AAT comb.		.83	.23	1.69	.83	.19	1.83	¹						
Lenient criterion														
FSQ	≥ 63	.97	.07	3.36	.97	.23	2.62	≤ 1	.86	.29	1.63	.86	.29	1.64
SAS	≥ 19	.93	.14	2.56	.93	.29	2.03	≤ 1	.90	.20	2.12	.90	.28	1.86
BAT	≤ 29 cm/s	.95	.14	2.73	.95	.25	2.32	≥ 41 cm/s	.93	.68	1.01	.93	.64	1.12
AAT RT	≥ 790 ms	.63	.44	0.48	.63	.23	1.07	¹						
AAT error	$\geq 1\%$.75	.54	0.57	.75	.41	0.90	¹						
AAT comb.		.89	.76	0.52	.89	.58	1.02	¹						

Notes: ¹Cutoff values for the AAT and the Neutral vs. Fake-low comparison were not computable because participants were unable to modify their AAT responses in the fake-low condition. ²For these tests no empirical threshold values existed near the derived criterion values.

the cutoff values for all measures and diagnostic criteria are listed with the corresponding d' values, hit rates, and false alarm rates. Here d' is reported, as it characterises the sensitivity for a fixed criterion appropriately.

Self-report measures. Average detection performance for detecting fake-high or fake-low results in the FSQ was $d_a = 3.46$ and $d_a = 1.70$, respectively; and $d_a = 2.70$ and $d_a = 2.09$ for the SAS. For all diagnostic criteria, performance for detecting fake-high results was nearly perfect for the FSQ, $d' \geq 3.36$, and lower but still very high for the SAS, $d' \geq 2.56$. This was different for identifying fake-low data: Here the SAS performed better, $d' \geq 2.12$, than the FSQ, $d' \geq 1.45$. Note that no cutoff values for the conservative criterion of the FSQ and the unbiased criterion of the SAS could be derived, as no empirical cutoff values near the derived ROC slopes existed.

Behavioural measure. Average detection performance was $d_a = 2.72$ and $d_a = 0.71$ for detecting fake-high and fake-low results, respectively. Noteworthy, setting the cutoff value to 0 cm/s (participants didn't move at all), already detected 30% of the fake-high results successfully without any false alarms.

Indirect measure. A two-step approach was used in analysing the AAT: First we used errors and median RTs separately to detect faked results; later we combined both to optimise detection performance. We used only trials with spider pictures to be pulled, as participants strategically faked on those trials. Notably, RTs and errors did not significantly differ between the fake-low and the neutral condition, showing that participants were unable to strategically change their responses in the fake-low condition. Therefore, detection performance and cutoff values were not determined for this condition.

The analyses revealed two easily detectable faking strategies: Some participants frequently pushed spider pictures away irrespective of instructions, others responded very slowly when they had to pull spider pictures closer. The first strategy was evident in the error rates of the fake-high condition, with an average detection performance of $d_a = 1.03$. Interestingly, detection performance varied greatly between the different diagnostic criteria, due to different variances in the neutral and fake-high error rates and resulting asymmetric ROC curves. Whereas performance was very high for the conservative criterion, $d' = 2.26$, it dropped to $d' = 0.57$ for the lenient criterion. The same was true for median RTs, assessing the second faking strategy: Performance for the conservative criterion was good, $d' = 1.70$, but dropped to $d' = 0.48$ for the lenient criterion, resulting in an average detection performance of $d_a = 0.50$.

To assess the performance of errors and RTs together, we determined for each diagnostic criterion, which error/RT data pairs were labelled as fakers by at least one of the cutoff values. This combination yielded a better performance, being most pronounced for the conservative criterion, $d' = 2.59$, and only minimal for the lenient criterion, $d' = 0.56$, showing that most participants adopted only one of the faking strategies.

Detection performance including independent data

Two aspects of the earlier detection analysis are potentially problematic: First, our sample contained few high or low spider-fearful participants, possibly inflating detection performance. Second, detection performance is assessed with the same data used to define the cutoff values, also possibly overestimating detection performance. To estimate detection performance in a more realistic context, we included data from high and low spider-fearful individuals of independent studies and recalculated the signal detection parameters (see Table 2). Note that hit rates stayed the same, because the amount of faked data did not change.

Self-report measures. For the FSQ and SAS we included data from Becker and Rinck (2004), Rinck et al. (2002), and Rinck and Becker (2007). For both questionnaires, this raised the false alarm rates of the fake-high condition by 12–14%, resulting in lower d' . Nevertheless, detection performance still was very good for both measures and all diagnostic criteria, $d' \geq 2.03$. For the fake-low condition, only the performance of the SAS was affected, with 8–9% higher false alarm rates, $d' \geq 1.72$.

Behavioural measure. For the BAT we included data from Becker and Rinck (2004), Lange, Heuer, Reinecke, Becker, and Rinck (2008), and Reinecke, Becker, and Rinck (2008). As for the questionnaires, false alarm rates in fake-high condition increased by 10–11%, resulting in lower, but still very high detection performance, $d' \geq 2.21$. For the fake-low condition, performance was worse for the conservative and unbiased criterion, $d' \geq 0.47$, but slightly higher for the lenient criterion, $d' = 1.12$.

Indirect measure. We included AAT RT and error data of three experiments by Rinck and Becker (2007). For RTs, errors, and the combination of both, false alarm rates were either unchanged or substantially lower for all diagnostic criteria, resulting in unchanged or higher detection performance (see Table 2).

DISCUSSION

In the present study, we investigated whether direct, indirect, and behavioural measures of fear of spiders could be faked successfully. For this, we first tested whether participants could change their test responses according to given faking instructions. We then determined how well faked responses could be identified by each measure and we provided cutoff values for three diagnostic contexts.

None of the measures was immune against faking, at least not under both faking instructions. As expected, participants could substantially change their scores on the direct measures. Sum scores of both questionnaires lay near the minimum value in the fake-low and near the maximum value in the fake-high condition, showing participants' good understanding of how the test items measure fear of spiders. For the BAT, participants adjusted their approach speed towards a spider corresponding to the instructions. Only for the AAT, fakeability depended on the instruction: Participants could change their responses in the fake-high condition, but seemed unable to do so in the fake-low condition. This is another example that indirect measures are not principally immune against faking.

All measures were capable of identifying fakers within the study sample successfully, although to a different degree: Both questionnaires and the BAT showed very good to near perfect detection performance for all diagnostic criteria, whereas the AAT showed comparable performance only for the conservative criterion. This indicates that the AAT might be suitable only when conservative detection of fakers with few false alarms is desired, whereas the direct and behavioural measures might be applicable over a wider range of situations. Noteworthy, using both AAT errors and RTs enhanced detection performance, reflecting that participants used different faking strategies.

Including data from truly high and low spider-fearful participants of other studies showed negative effects on detection performance of the direct and behavioural measures, but no effects or positive ones on AAT performance. This reflects a considerable overlap of really high or low spider-fearful data with faked data on the direct and behavioural measures: Extreme questionnaire scores and extreme approach behaviour do occur, and do not necessarily indicate faking. In contrast, there was little overlap between real and faked data on the AAT: In unfaked data, error rates above 10% or mean RTs larger than 1 s are extremely rare.

Our results indicate an inherent confound of fakeability and detectability: If participants can change their test scores easily to extreme values, these values can easily be detected. But although detection performance seems overestimated for some measures if we use only our sample, it stayed high

even when independent data were added, suggesting that the reported cutoff values are valid.

Some caution is warranted when generalising the results: The sample contained mainly female participants, as well as mainly young, computer-savvy students. Further, the faking conditions were always preceded by the neutral condition. Although we deliberately chose this order to ensure an unbiased neutral measurement and to give participants prior task experience (Steffens, 2004), it might have affected some measures, something that should be tested in future studies that avoid this order confound. Another important factor we did not control for is faking motivation, as individuals with higher faking motivation might produce data not as easily detectable. Although we encouraged participants to fake in an intelligent and not easily detected way, the impact of manipulating faking motivation by, for example, monetary incentives needs to be assessed.

Taken together, the question of whether the AAT is a viable alternative to direct or behavioural measures of spider fear has no simple answer. Regarding fakeability, the AAT seems preferable, as it was the only measure that participants were unable to fake under the fake-low condition. Considering detectability of fakers, the answer is complex. For conservative situations, where treating non-fakers correctly is more important, all measures discriminate well, even with independent data. For the unbiased and lenient criteria, the direct and behavioural measures outperform the AAT.

Obviously, it is unwarranted to condemn direct and behavioural measures while praising indirect measures as immune against faking. Instead, more studies are needed to identify the specific advantages and disadvantages of each task.

Manuscript received 14 March 2008

Revised manuscript received 25 November 2008

Manuscript accepted 26 November 2008

First published online day/month/year

REFERENCES

- Becker, E., & Rinck, M. (2004). Sensitivity and response bias in fear of spiders. *Cognition and Emotion*, 18(7), 961–976.
- De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology*, 43(6), 972–978.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327.
- Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, 27(4), 307–316.

- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, 66(1), 83–96.
- Klauer, K., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology*, 43(4), 648–655.
- Lange, W., Heuer, K., Reinecke, A., Becker, E. S., & Rinck, M. (2008). Inhibition of return is unimpressed by emotional cues. *Cognition and Emotion*, 22(8), 1433–1456.
- LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *The Journal of Applied Psychology*, 92(1), 1–16.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Muris, P., & Merckelbach, H. (1996). A comparison of two spider fear questionnaires. *Journal of Behavior Therapy and Experimental Psychiatry*, 27(3), 241–244.
- Reinecke, A., Becker, E. S., & Rinck, M. (2008). *Test-retest reliability and validity of three indirect tasks assessing implicit threat associations and behavioral response tendencies*. Manuscript submitted for publication.
- Rinck, M., & Becker, E. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry*, 38(2), 105–120.
- Rinck, M., Bundschuh, S., Engler, S., Muller, A., Wissmann, J., Ellwart, T., et al. (2002). Reliability and validity of German versions of three instruments measuring fear of spiders. *Diagnostica*, 48(3), 141–149.
- Schnabel, K., Banse, R., & Asendorpf, J. (2006). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: The Implicit Association Procedure (IAP). *Experimental Psychology*, 53(1), 69–76.
- Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology*, 59, 239–245.
- Steffens, M. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology*, 51(3), 165–179.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1(1), 1–26.
- Szymanski, J., & O'Donohue, W. (1995). Fear of spiders questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, 26(1), 31–34.
- Teige-Mocigemba, S., & Klauer, K. C. (2008). “Automatic” evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology*, 44(5), 1414–1417.
- Winer, B. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.